社会互动加工的认知特性及脑机制——第三人称的视角

程羽慧 袁祥勇 蒋毅

(中国科学院心理研究所，脑与认知科学国家重点实验室，脑科学与智能技术卓越创新中心，北京 100101)

(中国科学院大学心理学系，北京 100049)

(北京脑科学与类脑研究中心，北京 102206)

摘 要 在日常生活中，从“第三人称”的视角识别并理解他人的社会互动至关重要。这种社会互动加工具有两种认知特性：构形整体性和动作关联性；体现为一个由众多脑区共同参与的层级加工过程，主要包括个体知觉网络、动作观察网络和心智化网络。其中，后颞上沟等脑区在表征社会互动关系中起着关键作用。未来的研究需要结合多种技术手段进一步揭示社会互动加工的遗传特性和神经机制，并关注它在实际生活中的应用。

关键词 社会互动；第三人称视角；构形加工；动作关联性；脑机制

分类号 B842

1 引言

社会互动(social interaction)是指个体通过动作或语言表达某种交流的意图，而对方理解这种意图并做出恰当的回应，从而达成有意义的社会交流过程(Capozzi & Ristic, 2018; Hauser & Wood, 2010)。在日常生活中，社会互动无处不在，与朋友聊天、与舞伴跳舞、与对手格斗都属于不同形式的社会互动。研究社会互动有两种不同的视角：“第二人称”视角(second-person perspective)和“第三人称”视角(third-person perspective)。“第二人称”视角的研究一般通过创造实时的互动情境，考察被试与他人直接进行互动时的认知过程(Caruana, McArthur, Woolgar, & Brock, 2017; Schilbach et al., 2013)。与之相对，“第三人称”视角的研究主要关注被试作为旁观者，在观察他人进行社会互动时的认知过程(Quadflieg & Koldewyn, 2017)，包括对互动双方特征的知觉和意图的推断等。观察并理解他人的社会互动不仅对于人类的生存进化具有重要的意义，而且在我们的日常生活中起着重要的作用。它提供了个体观察学习的

收稿日期: 2020-05-14

* 国家自然科学基金(31830037, 31525011)、中国科学院战略性先导科技专项(XDB32010300)、中国科学院前沿科学重点研究计划项目(QYZDB-SSW-SMC030)和中央高校基本科研业务费专项资金资助。

通信作者：袁祥勇, E-mail: yuanxy@psych.ac.cn

机会，有助于个体更好地了解他人的个性特征和社会关系，从而加深对社会群体的理解，提高社会交往的技能(Matheson, Moore, & Akhtar, 2013; Milinski, 2016)。

近年来，越来越多的研究从“第三人称”视角来探究社会互动，这既可以揭示人类加工他人互动的认知神经机制，同时也为特殊人群的诊断鉴定和识别社会互动的人工智能算法提供了参考，因此具有重要的理论意义和应用价值。本文从“第三人称”视角出发，论述了社会互动加工的前提条件和认知特性，并在梳理了社会互动加工的认知神经机制的基础上，探讨了需要进一步研究的科学问题。因此，下文提到的社会互动特指“第三人称”视角的社会互动。

2 社会互动加工的前提条件

社会互动具有多种多样的呈现方式，既包括简短的视频(Sinke, Sorger, Goebel, & de Gelder, 2010)，静止的图片(Quadflieg, Gentile, & Rossion, 2015)，也包括将人物抽象成关节点的光点动画(Manera, Schouten, Becchio, Bara, & Verfaillie, 2010)和只具备生物运动模式的简单几何图形(Gao, Scholl, & McCarthy, 2012)。人们仅仅通过观察这些社会互动的刺激，就可以判断出双方是否在进行互动，具体在进行哪种互动(Manera, von der Luhe, Schilbach, Verfaillie, & Becchio, 2016)，进而可以推断他们是友好还是敌对(Wu, Hua, Yang, & Yin, 2018)，关系是否亲密(Costanzo & Archer, 1989)，或者彼此的社会地位是否相同等多种社会信息(Floyd & Erbert, 2003; Mast & Hall, 2004)。

不管以哪种方式呈现，只有同时满足空间、时间及意义维度上的前提条件，两个个体才会被知觉为正在进行社会互动。具体而言，在空间维度上，两个个体的空间距离和身体朝向都会影响观察者对社会互动的感知。最近的研究发现两个个体的距离越近，身体越是朝向对方，他们被判断为正在互动的可能性就越大(Hartmann et al., 2019; Zhou, Han, Liang, Hu, & Kuai, 2019)。在时间维度上，当两个个体的动作具有连续性时，即小人 A 的动作完成后小人 B 做出及时的回应，他们更可能被认为在进行社会互动。而随着双方动作时间间隔的增大，小人 A 的动作对小人 B 的预测作用也会相应变差(Manera, Schouten, Verfaillie, & Becchio, 2013)，这意味着观察者对双方互动的知觉也随之变差。在意义维度上，两个个体的动作只有形成符合逻辑的联系，才有可能被认为存在社会互动关系。例如，当小人 A 发出“站起来”的动作指令，小人 B 从蹲下的姿态站了起来，两个小人会被知觉为正在互动；如果 A 发出

同样的指令，而 B 从始至终一直在扫地，他们就会被认为是相互独立的个体，彼此没有互动(Manera et al., 2015)。综上所述，距离相近、身体相向、时间连续、意义合理都是感知到社会互动的前提条件。绝大多数研究者会基于上述前提条件之一，生成背对背的身体姿态，打乱互动双方动作的时间结构，或者匹配两个没有逻辑关联的动作，作为控制实验的非社会互动刺激。然而，在真实的社会情境中，上述因素往往共同影响着对社会互动的判断，但只有极少数研究同时考察了多种因素在社会互动知觉中的作用(Hartmann et al., 2019)，未来可从整合的视角考察多种因素的共同作用方式。

3 社会互动加工的认知特性

两个互动的个体一旦被第三方知觉为正在进行社会互动，他们在空间上就会以整体构形(global configuration)的方式得到加工，而他们的动作则会形成彼此关联且有意义的时间结构得到表征。这种构形的整体性和动作的关联性既是社会互动加工的认知特性，也是社会互动具有加工优势的潜在原因(Fedorov, Chang, Giese, Bulthoff, & de la Rosa, 2018; Neri, Luu, & Levi, 2006; Papeo, Goupil, & Soto-Faraco, 2019; Vestner, Tipper, Hartley, Over, & Rueschemeyer, 2019)。下文将详细阐述支持这两种认知特性的证据，以及它们所带来加工优势的具体表现。

3.1 整体性构形加工

社会互动的整体性构形加工是指人们对两个互动的个体所形成的空间结构(面对面的身体朝向)进行整体表征，即互动双方被当作一个结构化的功能整体而非两个独立的个体得到加工和存储(Ding, Gao, & Shen, 2017; Papeo et al., 2019; Papeo, Stein, & Soto-Faraco, 2017; Vestner et al., 2019)。假设社会互动存在整体性构形加工，那么互动双方作为一个整体必然会影响空间注意的分配(Shen, Yin, Ding, Shui, & Zhou, 2016; Yin, Xu, Duan, & Shen, 2018)，表现出经典的客体注意效应(object-based attentional effect)，即注意资源会自动分配在同一客体的不同部分，对它们的反应要快于对其它客体的反应。在社会互动情境中，当线索呈现在一个小人身上时，被试不仅对同一小人身上的目标反应更快，对与他互动小人身上的目标反应也快于其他小人身上的目标；而此效应在非互动的情景下消失(Ji, Yin, Huang, & Ding, in press)。不仅如此，社会互动的整体性构形加工还直接体现在由于互动双方被知觉为一个整体，他们的主观距离比实际距离更近(Vestner et al., 2019)。

在整体性构形加工中，互动双方的空间结构以整体的方式得到表征。相比于对局部空间信息的逐步加工，这种整体加工的方式在一定程度上减少了认知负荷，从而使互动刺激比非互动刺激具有加工优势(Papeo et al., 2019; Papeo et al., 2017; Vestner, Gray, & Cook., 2020)。首先，整体性构形加工有助于互动刺激快速通达意识，被试识别被掩蔽到阈下的两个面对面互动小人的正确率显著高于被掩蔽的背对背非互动小人；而对于倒置的刺激，被试则不会表现出这种识别优势(Papeo et al., 2017)。其次，整体性构形加工也有助于快速搜索互动刺激，表现出搜索不对称性，即从一群背对背小人中搜索面对面小人比从一群面对面小人中搜索背对背小人的正确率更高，速度更快(Papeo et al., 2019)。最后，整体性构形加工还有利于记忆的存储和提取。不论在短时记忆任务还是长时记忆的任务中，被试记忆互动刺激的正确率都高于非互动刺激(Ding et al., 2017; Vestner et al.,2019)。以上这些研究不仅证明了整体性构形加工所带来的加工优势，也再次印证了社会互动加工中存在整体性构形表征。但是，关于整体性构形加工是特异于社会互动刺激还是也存在于其他面对面的客体中，目前还尚未得出一致的结论(Papeo & Abassi, 2019; Vestner et al., 2020)。假如整体性构形加工特异于社会互动刺激，是否适用于所有类型(如合作或竞争)的社会互动也有待进一步的研究(Yin et al., 2013)。

3.2 关联性动作表征

社会互动不仅在空间上作为一个整体被加工，同时也是一个动态连续的时间过程。在社会互动中，不论是社交发起者传递交流意图，回应者做出回应(例如，A 扔出东西，B 接住)，还是互动双方通过默契的配合，完成协调性的联合动作(joint action, 如双人舞蹈; Marsh, Richardson, & Schmidt, 2009; Pesquita, Whitwell, & Enns, 2018)，双方的动作都是有联系的，它们共同组成了一个有意义的、连续的时间结构，这使得互动动作以相互关联的形式在大脑中被表征。有研究者使用了两对互动动作——其中“拿来”回应“给予”，“接住”回应“扔出”——来检验观察一个动作是否能够自动触发与它相关联的动作表征。他们发现，给被试重复呈现“拿来”的动作，会让他们更倾向于将既像“拿来”又像“接住”的两可动作知觉成“接住”的动作；更为关键的是，即使重复呈现“给予”的动作后，被试也会把这个两可动作知觉成“接住”的动作。可见，尽管“给予”与“拿来”的动作存在很大的视觉差异，却可以引发相同的适应结果。这种交叉适应效应(cross-adaptation effects)充分证明了互动动作存在关联性的表征

(Fedorov et al., 2018)。此外，关联性动作表征的证据还体现在被试对互动动作的时间感知更短(Liu, Yuan, Chen, Jiang, & Zhou, 2018)，动作流畅性更高(Peng, Ichien, & Lu, 2020)。

关联性动作表征使得人类对他人互动行为的感知不仅基于当前的视觉输入，还可以依赖先前的知识经验和动作之间连续的时间结构，预测接下来可能的动作，从而在不确定性较高的情况下，更好地识别互动动作(Manera, Becchio, Schouten, Bara, & Verfaillie, 2011; Manera, Giudice, Bara, Verfaillie, & Becchio, 2011; Neri et al., 2006; von der Luhe et al., 2016)。Neri 和他的同事(2006)最早证明了这种促进效应。他们给被试先后呈现两个被噪音点掩蔽的光点动画，其中一个动画包含两个互动或非互动的小人，另一个动画则只包含其中的一个小人，结果发现当双人动画是时间结构完整的互动动画时，被试判断双人动画的正确率更高。随后的研究则进一步证实，被试可以利用社交发起者的动作来预测回应者的动作，具体表现为当两个小人 A 和 B 中的回应者 B 被噪音掩蔽时，若小人 A 和 B 具有互动关系，被试从噪音中辨别小人 B 会更容易(Manera, Becchio, et al., 2011; Manera, Giudice, et al., 2011; Manera et al., 2013)。除了噪音掩蔽范式，这种基于关联性动作的预测效应在双眼竞争(binocular rivalry)范式中也得到了验证(Su, Van Boxtel, & Lu, 2016)。尽管选取了不同的实验范式和多样的刺激形式，上述研究都得出了较为一致的结论：关联性动作表征可以促进互动动作的识别。不过目前的研究主要以整个身体的动作为主，较少关注手势、脚部动作等局部肢体动作(Zaini, Fawcett, White, & Newman, 2013)；而在跳舞等某些特定情景中，局部肢体动作也是重要的信息传递载体，未来的研究可以应用更具生态效度的刺激来探索局部肢体动作的关联性表征在社会互动中的作用。

4 社会互动加工的脑机制

作为高级认知活动，社会互动加工是一个从识别双方身体和面孔到理解互动的动作，从而推测其交流意图的层级加工过程。这个过程可以简单分为三个阶段，分别对应于三个脑网络(见图 1)：个体感知网络(person perception network)、动作观察网络(action observation network)和心智化网络(mentalizing network)，详见 Quadflieg 和 Koldewyn(2017)的综述。个体感知网络负责对互动者的身体和面孔进行视觉分析，主要包括特异于面孔加工的梭状回(fusiform face area, FFA)、肢体加工的外纹状皮层身体区(extrastriate body area, EBA)、动态面孔和生物运动加工的后颞上沟(posterior superior temporal sulcus, pSTS; Quadflieg et al., 2015)。

动作观察网络属于镜像神经元系统(mirror neuron system, MNS)，主要参与对他人动作的快速识别及意图的准确理解，由前运动皮层(premotor cortex, PMC)、额下回(inferior frontal gyri, IFG)和顶下小叶(inferior parietal lobule, IPL)等组成(Caspers, Zilles, Laird, & Eickhoff, 2010)。心智化网络是心理理论(theory of mind)的神经基础，主要负责判断他人抽象的心理状态(如欲望、动机、意图、观念等)和推测复杂的社会属性(如社会规范，亲属关系，情绪状态等; Keysers & Gazzola, 2007)，它主要包括背/腹内侧前额叶皮层(dorsal/ventral medial prefrontal cortex, dmPFC/vmPFC)、颞顶联合区(temporoparietal junction, TPJ)、楔前叶(precuneus, PrC)等脑区(Schurz, Radua, Aichhorn, Richlan, & Perner, 2014)。

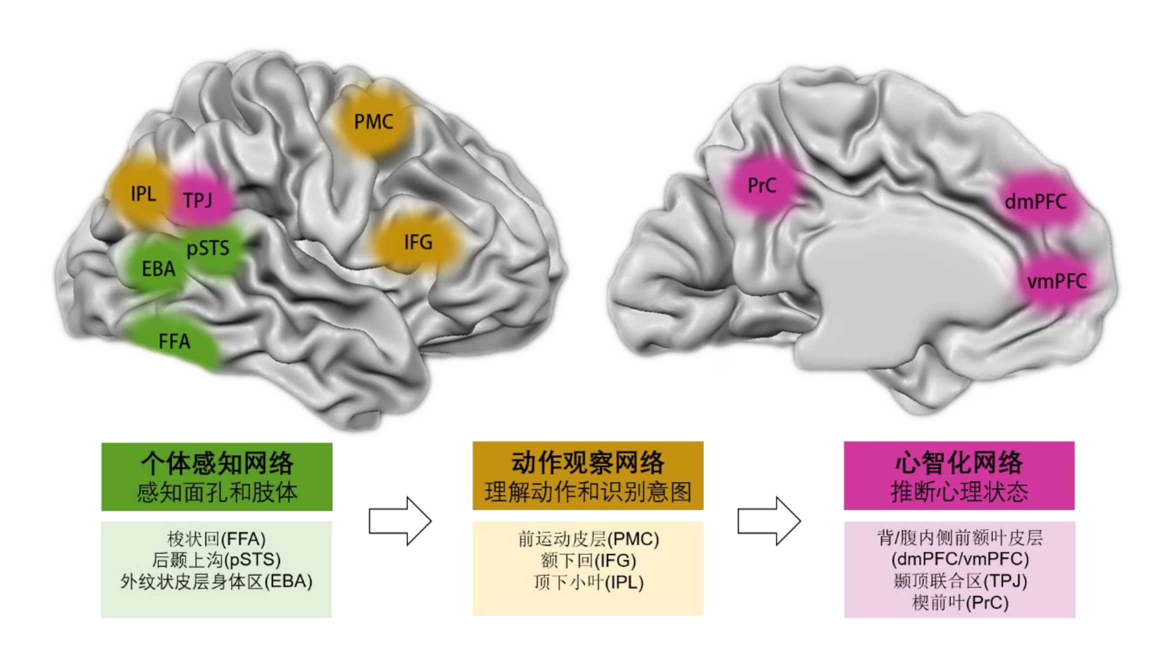


图 1 社会互动加工过程的脑网络。各个脑区分别对应着下方相同颜色的脑网络。

4.1 与社会互动认知特性相关的脑区

在揭示社会互动加工的脑网络时，研究者主要关注被试观看互动刺激所激活的脑区，但随着行为范式和数据处理方法的发展，更多的研究者也开始寻找与社会互动认知特性加工相关的脑区。首先，是否存在特定的脑区支持社会互动的整体构形加工？已有证据提示，属于个体感知网络的 EBA 脑区，对身体的姿态信息敏感，并基于此对动作进行编码(Downing, Peelen, Wiggett, & Tew, 2006)，因此可能是与社会互动整体构形加工相关的脑区。为了证明这个推断，研究者利用多体素模式分析技术(multivoxel pattern analysis, MVPA)技术，训练分类器根据 pSTS 或 EBA 的激活模式来判断被试观看互动视频的类型(争论、庆祝或大笑)，然

后分别测试辨别两人互动动作以及其中单人动作的成绩。测试结果表明，pSTS 能够分辨互动动作和单人动作，但是两者的正确率没有显著差异；而 EBA 区分互动动作的正确率要显著高于区分单人动作的正确率，这说明 EBA 除了加工局部单人动作，也提取了社会互动独特的整体信息(Walbrin & Koldewyn, 2019)。随后的研究进一步表明 EBA 的整体编码只限于面对面的小人，也就是说，在训练分类器对单个小人的动作进行区分后，当这个小人与另一个小人面对面而非背对背时，通过 EBA 的激活模式识别出此小人动作的正确率更高。即当两个小人的身体相向从而形成社会互动关系时，EBA 编码单个小人身体的能力也得到增强(Abassi & Papeo, 2020)。上述结果都证明了 EBA 在社会互动的整体性构形加工中发挥了重要的作用(Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019)。

其次，研究者也关注是否存在某些脑区表征社会互动的关联性动作。由于关联性动作表征涉及到互动动作的时间连续和意义关联，因此可能与动作观察网络和心智化网络有关。在社会互动中，关联性动作相比于非关联性动作(用一个小人的镜像动作替代与他互动的另一人的动作)更强的激活了包括 IFG(延伸到 PMC)，pSTS 和左侧 IPL 等在内的动作观察网络(Georgescu et al., 2014)。此外，Zillekens 等人(2019)利用前人开发的噪音掩蔽范式(Manera, Giudice, et al., 2011)初步考察了社会互动中预测他人动作的神经机制。他们通过操纵小人 A 是否向小人 B 传递交流的意图，将动画分为互动与非互动动画；同时将小人 B 用噪音掩蔽，要求被试判断小人 B 是否存在。结果发现，无论小人 B 是否存在，互动动画比非互动动画都显著的增强了额上回(superior frontal gyrus, SFG)，相反则激活了顶上小叶(superior parietal lobule, SPL)、颞下回(inferior temporal gyri, ITG)等属于动作观察网络的脑区；上述结果表明，在互动情景下大脑能够根据小人 A 的动作形成对小人 B 动作的预期，这一定程度上减少了执行控制的需求进而表现出动作观察网络激活水平的降低。当违反预期时，即在互动动画中 B 不存在或在非互动动画中 B 存在时，右侧后内侧额回(right posterior medial frontal gyrus, rPMFG)、双侧楔前叶、左侧小脑和左侧梭状回都被更强激活，这些脑区可能与社会互动中预测偏差有关。此外，研究还发现杏仁核的激活程度可以预测被试辨别 B 的能力，并且在互动情境条件下杏仁核与 mPFC 的功能连接也得到增强，这反映了对互动动作的预期涉及较复杂的社会推断，可能需要整合心智化网络的功能(Zillekens et al., 2019)。

上述研究发现了表征社会互动认知特性所对应的脑区，为揭示整体性构形加工与关联性动作表征的神经机制提供了初步的方向。但许多重要的问题还有待回答，如依据互动发起者

的动作激活模式是否可以解码出回应者的动作。另外，研究还亟需建立行为任务表现与脑成像结果的联系从而更准确地揭示行为与脑功能的关系，如 EBA 的激活是否可以预测整体性构形加工任务的成绩并且反映知觉社会互动的个体差异。

4.2 社会互动加工的特异性脑区

尽管社会互动加工在不同的阶段激活了众多不同的脑区，但有些脑区可能仅仅参与了单个人的身体感知和动作理解，因此，研究者试图通过更精细的实验设计分离出真正能表征互动关系的特异性脑区。在一个以猕猴为被试的实验中，研究者利用功能磁共振成像技术(functional magnetic resonance imaging, fMRI)考察猕猴在观看不同互动视频时全脑的激活情况(Sliwa & Freiwald, 2017)，结果发现只有内侧前额叶以及 IPL 的部分脑区(对应于人类的心智化网络)，特异于加工猴子与猴子之间的社会互动，而对猴子与物体互动、物体与物体互动的视频都不敏感。这说明猕猴大脑中存在表征社会互动关系的特异性脑区。但是，人类大脑中是否也存在类似的脑区？研究者把目光聚集到了 pSTS，它不仅被认为是个体感知网络的脑区，也经常被看作动作观察网络和心智化网络的脑区(Deen, Koldewyn, Kanwisher, & Saxe, 2015; Yang, Rosenblau, Keifer, & Pelphrey, 2015)，并且它对多种类型的社会刺激都有反应，被称为社会大脑的“中枢”(Lahnakoski et al., 2012)。为了进一步排除无关变量对结果的影响，Isik 等人(2017)简化了社会互动刺激的物理属性，只保留了它们的互动关系。他们发现，由简单几何图形所组成的互动动画(两个图形相互阻碍或者帮助)比物理互动(两个图形相撞)或单人动画更强显著激活了 pSTS，而且研究者利用 MVPA 技术，通过 pSTS 的激活模式可以区分出互动关系是相互阻碍还是帮助。该实验表明 pSTS 能够表征社会互动关系(类似结果见 Walbrin, Downing, & Koldewyn, 2018)，可能是加工社会互动的特异性脑区(Isik, Koldewyn, Beeler, & Kanwisher, 2017)。但值得注意的是，在该研究中 pSTS 只是对社会互动反应最强的一个区域，它的边界也延伸到了附近表征心理理论的 TPJ 区域和功能定位的动态面孔脑区。综上，大脑中可能存在真正表征社会互动关系的特异性脑区——心智化网络及附近的 pSTS(Isik et al., 2017)，这些脑区对应于理解双方意图及心理状态的加工阶段。

4.3 社会互动加工脑区的功能连接

近年来，一些研究者从层级加工的角度出发，进一步探讨不同脑区在整个加工过程中的信息传递和功能连接。一项脑成像元分析的研究指出，在社会互动加工中，从识别个体的动作到解析单个动作和互动动作的意图，再到推断动作意图背后的心理状态，主要激活了从颞叶中后部皮层经 pSTS 传向 TPJ 的通路(Arioli & Canessa, 2019)。其中，pSTS 很可能是社会互动加工过程中的重要节点，起着对信息进行分类筛选并向其它脑区传输的作用(Sliwa & Freiwald, 2017)。因此，pSTS 与其它脑区的功能连接也可能会受到社会互动类型的调节。在需要理解动作的共同意图时，比如两个人合作搬箱子，pSTS 增强了与动作观察网络中的顶上小叶(superior parietal lobule, SPL)和腹侧 PMC 的连接，而在需要理解情绪的互动中，比如两个人微笑着牵手，pSTS 则增强了与心智化网络中 vmPFC 的连接(Arioli et al., 2018)。目前有关功能连接的研究只是刚刚起步，未来需要更多的数据来全面刻画不同阶段和不同类型的社会互动加工所依赖的功能连接模式，并揭示关键节点如 pSTS 在社会互动信息传递中的作用。

5 总结与展望

在社会环境中，他人的社会互动为我们提供了丰富的社会信息，所以检测和识别社会互动信息是十分重要的。这种重要性使得社会互动同面孔和生物运动一样，人类对其具有较高的敏感性，并表现出对社会互动的加工优势：相比于非社会互动刺激，会优先加工社会互动刺激；识别与记忆社会互动刺激的能力也更强。人类对社会互动整体性构性加工和关联性动作表征是构成其加工优势的深层原因。社会互动是一个层级加工过程，所涉及的脑网络包括个体感知网络、动作观察网络和心智化网络(Quadflieg & Koldewyn, 2017)。对于社会互动加工的认知特性来说，初步的研究显示 EBA 参与社会互动的整体性构形加工，而动作观察网络和心智化网络的一些脑区则与关联性动作表征有关。尽管存在某些特定的脑区对社会互动关系更为敏感，但是各个脑区在不同加工阶段的信息传递和相应的功能连接对于完成社会互动加工也同样重要(Arioli & Canessa, 2019)。总结来看，尽管研究者对社会互动加工进行了多方面的探索，但还有一些有待解决的问题。

第一，人类能够快速探测到社会互动信号，正确解读他人的动作含义，但这种能力是否来自先天遗传还未有定论。研究显示，5~14 个月的婴儿更加偏好社会互动的信息，当屏幕两侧同时呈现互动和非互动的动画时，他们对互动动画的注视时间更长(Galazka, Roche,

Nystrom, & Falck-Ytter, 2014)，而且在两个个体动态追逐的动画中，他们的目光注视“追”的个体时间显著长于“逃”的个体和不参与互动的个体(Galazka & Nystrom, 2016)。这说明不到 1 岁的婴儿已经对环境中的互动信息及引发互动的个体表现出更多的兴趣，表明这种能力可能是人类在进化过程中通过自然选择而保留下来，因此具有一定的遗传性。未来可通过行为遗传学的方法，比如双生子实验(Wang et al., 2018)，对社会互动的遗传特性提供更为直接的证据。

第二，以往的很多研究都指出，人类对诸如面孔、生物运动等传递出大量生物性和社会性信息的刺激都具有整体加工优势，即加工依赖于各个部分之间的空间结构关系并表现出经典的倒置效应(Farah, Tanaka, & Drain, 1995; Reed, Stone, Bozova, & Tanaka, 2003)。相比于精细的轮廓线等具有较高空间频率的信息，较为粗略的整体结构体现在刺激低空间频率的信息中，且加工速度快于高频信息，因此这种整体加工优势通常被认为只存在于刺激的低频信息(Goffaux et al., 2011)。与面孔和生物运动类似，社会互动也以整体构形的方式被加工，并且传递出重要的社会性信息，具有进化意义。所以社会互动的整体构形可能特异于互动刺激的低频成分，从而表现出相对于非互动刺激的加工优势。但是，上述推断是否正确还需要进一步的实验进行探究。

第三，相关脑机制的研究已经表明社会互动加工是一个层级加工过程。尽管现有的研究揭示了动作观察网络、心智化网络和 pSTS 在社会互动加工中的作用(Arioli & Canessa, 2019; Sliwa & Freiwald, 2017)，但无论是从社会互动认知特性的脑机制，特异性脑区，还是脑区之间功能连接的角度，仍有许多问题值得探索。比如，在预测他人动作时，相应脑区的功能连接是否随互动的进程而动态变化；pSTS 是否具有筛选和传递社会互动信息的功能；不同类型社会互动的脑机制有多大的异同(Canessa et al., 2012; Sinke et al., 2010)，是否依赖于不同的功能连接模式(Arioli et al., 2018)等等。最后，虽然本文主要关注“第三人称”社会互动的加工，但随着多人脑神经活动同步记录技术(hyperscanning)的发展，人与人之间实时互动的脑机制正逐渐成为 “第二人称”社会互动研究的重点(Kingsbury & Hong, 2020)。然而，目前直接对比两种视角下社会互动加工的研究还较少，未来可结合多种技术手段探索它们在脑机制上的异同。

最后，目前的研究多集中于对社会互动的认知及神经机制的探讨，未来可以多发掘社会互动加工的应用价值。已有多项证据表明某些疾病患者的社会互动加工能力会受到损害(Centelles, Assaiante, Etchegoyhen, Bouvard, & Schmitz, 2013; Kuschefski, Falter-Wagner, Bente, Vogeley, & Georgescu, 2019; Okruszek et al., 2015; Walter et al., 2009)，例如，高功能自闭症患者无法利用动作关联性预测他人的动作(von der Luhe et al., 2016)，而精神分裂症患者虽然表征关联性动作的能力没有受到损伤，但却表现出对社会互动中生物运动识别能力的下降(Okruszek, Piejka, Wysokinski, Szczepocka, & Manera, 2018, 2019)。未来可利用更具有生态效度的虚拟现实技术来探索这些疾病患者加工社会互动的认知特性，并为相关疾病的诊断及干预提供新的思路。此外，利用人工智能识别现实场景中人与人的互动行为，例如在公共场所对可疑群体的检测，也具有重要、广阔的应用前景。

参考文献

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *The Journal of Neuroscience, 40*(4), 852–863.

Arioli, M., & Canessa, N. (2019). Neural processing of social interaction: Coordinate‐based meta‐analytic evidence from human neuroimaging studies. *Human Brain Mapping*, *40*(13), 3712–3737.

Arioli, M., Perani, D., Cappa, S., Proverbio, A. M., Zani, A., Falini, A., & Canessa, N. (2018). Affective and cooperative social interactions modulate effective connectivity within and between the mirror and mentalizing systems. *Human Brain Mapping, 39*(3), 1412–1427.

Canessa, N., Alemanno, F., Riva, F., Zani, A., Proverbio, A. M., Mannara, N., . . . Cappa, S. F. (2012). The neural bases of social intention understanding: The role of interaction goals. *PLoS ONE, 7*(7), e42347.

Capozzi, F., & Ristic, J. (2018). How attention gates social interactions. *Annals of the New York Academy of Sciences, 1426*(1), 179–198.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews, 74,* 115–125.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage, 50*(3), 1148–1167.

Centelles, L., Assaiante, C., Etchegoyhen, K., Bouvard, M., & Schmitz, C. (2013). From action to interaction: Exploring the contribution of body motion cues to social understanding in typical development and in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(5), 1140–1150.

Costanzo, M., & Archer, D. (1989). Interperting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior, 13*(4), 225–245.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex, 25*(11), 4596–4609.

Ding, X., Gao, Z., & Shen, M. (2017). Two equals one: Two human actions during social interaction are grouped as one unit in working memory. *Psychological Science, 28*(9), 1311–1320.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience, 1*(1), 52–62.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628–634.

Fedorov, L. A., Chang, D. S., Giese, M. A., Bulthoff, H. H., & de la Rosa, S. (2018). Adaptation aftereffects reveal representations for encoding of contingent social actions. *Proceedings of the National Academy of Sciences, 115*(29), 7515–7520.

Floyd, K., & Erbert, L. A. (2003). Relational message interpretations of nonverbal matching behavior: An application of the social meaning model. *The Journal of Social Psychology, 143*(5), 581–597.

Galazka, M., & Nystrom, P. (2016). Infants' preference for individual agents within chasing interactions. *Journal of Experimental Child Psychology, 147*, 53–70.

Galazka, M. A., Roche, L., Nystrom, P., & Falck-Ytter, T. (2014). Human infants detect other people's interactions based on complex patterns of kinematic information. *PLoS ONE, 9*(11), e112432.

Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *The Journal of Neuroscience, 32*(41), 14276–14280.

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions. *Human Brain Mapping, 35*(4), 1362–1378.

Goffaux, V., Peters, J. C., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, 21*(2), 467–476.

Hartmann, M., Mavrolampados, A., Allingham, E., Carlson, E., Burger, B., & Toiviainen, P. (2019). Kinematics of perceived dyadic coordination in dance. *Scientific Reports, 9*(1), 15594.

Hauser, M., & Wood, J. (2010). Evolving the capacity to understand actions, intentions, and goals. *Annual Review of Psychology, 61*(1), 303–324.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences, 114*(43), E9145–E9152.

Ji, H., Yin, J., Huang, Y., & Ding, X. (in press). Selective attention operates on the group level for interactive biological motion. *Journal of Experimental Psychology: Human Perception and Performance*.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences*, 11(5), 194–196.

Kingsbury, L., & Hong, W. (2020). A multi-brain framework for social interaction. *Trends in neurosciences*, 43(9), 651–666.

Kuschefski, M., Falter-Wagner, C. M., Bente, G., Vogeley, K., & Georgescu, A. L. (2019). Inferring power and dominance from dyadic nonverbal interactions in autism spectrum disorder. *Autism Research, 12*(3), 505–516.

Lahnakoski, J. M., Glerean, E., Salmi, J., Jaaskelainen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience, 6*(8), 233.

Liu, R., Yuan, X., Chen, K., Jiang, Y., & Zhou, W. (2018). Perception of social interaction compresses subjective duration in an oxytocin-dependent manner. *Elife, 7*, e32100.

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE, 6*(1), e14594.

Manera, V., Giudice, M., Bara, B. G., Verfaillie, K., & Becchio, C. (2011). The second-agent effect: Communicative gestures increase the likelihood of perceiving a second agent. *PLoS ONE, 6*(7), e22650.

Manera, V., Iani, F., Bourgeois, J., Haman, M., Okruszek, L. P., Rivera, S. M., . . . Becchio, C. (2015). The Multilingual CID-5: A new tool to study the perception of communicative interactions in different languages. *Frontiers in Psychology, 6*, 1724.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior Research Methods, 42*(1), 168–178.

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time will show: Real time predictions during interpersonal action perception. *PLoS ONE, 8*(1), e54949.

Manera, V., von der Luhe, T., Schilbach, L., Verfaillie, K., & Becchio, C. (2016). Communicative interactions in point-light displays: Choosing among multiple response alternatives. *Behavior Research Methods, 48*(4), 1580–1590.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*(2), 320–339.

Mast, M. S., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior, 28*(3), 145–165.

Matheson, H., Moore, C., & Akhtar, N. (2013). The development of social learning in interactive and observational contexts. *Journal of Experimental Child Psychology, 114*(2), 161–172.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1687), 20150100.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience, 9*(9), 1186–1192.

Okruszek, L., Haman, M., Kalinowski, K., Talarowska, M., Becchio, C., & Manera, V. (2015). Impaired recognition of communicative interactions from biological motion in schizophrenia. *PLoS ONE, 10*(2), e0116793.

Okruszek, L., Piejka, A., Wysokinski, A., Szczepocka, E., & Manera, V. (2018). Biological motion sensitivity, but not interpersonal predictive coding is impaired in schizophrenia. *Journal of Abnormal Psychology, 127*(3), 305–313.

Okruszek, L., Piejka, A., Wysokinski, A., Szczepocka, E., & Manera, V. (2019). The second agent effect: Interpersonal predictive coding in people with schizophrenia. *Social Neuroscience, 14*(2), 208–213.

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human-human interactions. *Journal of Experimental Psychology: Human Perception and Performance, 45*(7), 877–888.

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science, 30*(10), 1483–1496.

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science, 28*(3), 369-379.

Peng, Y., Ichien, N., & Lu, H. (2020). Causal actions enhance perception of continuous body movements. *Cognition, 194*, 104060.

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin & Review, 25*(5), 1751–1769.

Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex, 70*, 5–20.

Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences, 1396*(1), 166–182.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science, 14*(4), 302–308.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393–414.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9–34.

Shen, M., Yin, J., Ding, X., Shui, R., & Zhou, J. (2016). Deployment of attention on handshakes. *Frontiers in Psychology, 7*, 681.

Sinke, C. B., Sorger, B., Goebel, R., & de Gelder, B. (2010). Tease or threat? Judging social interactions from bodily expressions. *Neuroimage, 49*(2), 1717–1727.

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science, 356*(6339), 745–745.

Su, J., Van Boxtel, J. J., & Lu, H. (2016). Social interactions receive priority to conscious perception. *PLoS ONE, 11*(8), e0160468.

Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition, 200*, 104270.

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General, 148*(7), 1251–1268.

von der Luhe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B Biological Sciences, 371*(1693).

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia, 112*, 31–39.

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *Neuroimage, 198*, 296–302.

Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R. B., Erk, S., & Bara, B. G. (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: An fMRI study. *Social Cognitive and Affective Neuroscience, 4*(2), 166–176.

Wang, Y., Wang, L., Xu, Q., Liu, D., Chen, L., Troje, N. F., . . . Jiang, Y. (2018). Heritable aspects of biological motion perception and its covariation with autistic traits. *Proceedings of the National Academy of Sciences, 115*(8), 1937–1942.

Wu, X., Hua, R., Yang, Z., & Yin, J. (2018). The influence of intention and outcome on evaluations of social interaction. *Acta Psychologica, 182*, 75–81.

Yang, D. Y., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews, 51*, 263–275.

Yin, J., Ding, X., Zhou, J., Shui, R., Li, X., & Shen, M. (2013). Social grouping: Perceptual grouping of objects by cooperative but not competitive relationships in dynamic chase. *Cognition, 129*(1), 194–204.

Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-based attention on social units: Visual selection of hands performing a social interaction. *Psychological Science, 29*(7), 1040–1048.

Zaini, H., Fawcett, J. M., White, N. C., & Newman, A. J. (2013). Communicative and noncommunicative point-light actions featuring high-resolution representation of the hands and fingers. *Behavior Research Methods, 45*(2), 319–328.

Zhou, C., Han, M., Liang, Q., Hu, Y. F., & Kuai, S. G. (2019). A social interaction field model accurately identifies static and dynamic social groupings. *Nature Human Behaviour, 3*(8), 847–855.

Zillekens, I. C., Brandi, M.L., Lahnakoski, J. M., Koul, A., Manera, V., Becchio, C., & Schilbach, L. (2019). Increased functional coupling of the left amygdala and medial prefrontal cortex during the perception of communicative point-light stimuli. *Social Cognitive and Affective Neuroscience, 14*(1), 97–107.

The cognitive characteristics of and the brain mechanisms underlying social interaction processing from a third-person perspective

CHENG Yuhui; YUAN Xiangyong; JIANG Yi

(State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China)

(Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China)

(Chinese Institute for Brain Research, Beijing, 102206, China)

Abstract: The ability to recognize and decipher social interaction of others from a third-person perspective is critical for our daily life. There are two cognitive characteristics accounting for the processing of social interaction: configural integrity and action contingency. A hierarchical neural basis underpins social interaction processing in which the person perception network, the action observation network and the mentalizing network are concurrently engaged, wherein the posterior superior temporal sulcus plays a crucial role. Future research needs to explore the heredity of social interaction, to elucidate its underlying cognitive and neural mechanism by combining various technological methods, and to focus on its application in real life.

Key words: social interaction; third-person perspective; configural processing; action contingency; brain mechanism